



## King's Research Portal

DOI:

[10.1016/j.csda.2015.06.007](https://doi.org/10.1016/j.csda.2015.06.007)

*Document Version*

Peer reviewed version

[Link to publication record in King's Research Portal](#)

*Citation for published version (APA):*

Ryan, E. G., Drovandi, C., & Pettitt, A. (2015). Simulation-based fully Bayesian experimental design for mixed effects models. *COMPUTATIONAL STATISTICS AND DATA ANALYSIS*, 92, 26-39.

<https://doi.org/10.1016/j.csda.2015.06.007>

### **Citing this paper**

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### **General rights**

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### **Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Simulation-based Fully Bayesian Experimental Design for Mixed Effects Models

Elizabeth G. Ryan<sup>a,b,\*</sup>, Christopher C. Drovandi<sup>a,c</sup>, Anthony N. Pettitt<sup>a,c</sup>

<sup>a</sup>*Mathematical Sciences, Queensland University of Technology, Brisbane, Australia, 4000*

<sup>b</sup>*Biostatistics Department, Institute of Psychiatry, Psychology and Neuroscience, King's College London, United Kingdom, SE5 8AF*

<sup>c</sup>*ARC Centre of Excellence for Mathematical and Statistical Frontiers, Queensland University of Technology, Brisbane, Australia, 4000*

---

## Abstract

Bayesian inference has commonly been performed on nonlinear mixed effects models. However, there is a lack of research into performing Bayesian optimal design for nonlinear mixed effects models, especially those that require searches to be performed over several design variables. This is likely due to the fact that it is much more computationally intensive to perform optimal experimental design for nonlinear mixed effects models than it is to perform inference in the Bayesian framework. Fully Bayesian experimental designs for nonlinear mixed effects models are presented, which involve the use of simulation-based optimal design methods to search over both continuous and discrete design spaces. The design problem is to determine the optimal number of subjects and samples per subject, as well as the (near) optimal urine sampling times for a population pharmacokinetic study in horses, so that the population pharmacokinetic parameters can be precisely estimated, subject to cost constraints. The optimal sampling strategies, in terms of the number of subjects and the number of samples per subject, were found to be substantially different between the examples considered in this work, which highlights the fact that the designs are rather problem-dependent and can be addressed using the methods presented.

---

\*Corresponding Address: Elizabeth Ryan, Biostatistics Department, Institute of Psychiatry, Psychology and Neuroscience, King's College London, United Kingdom SE5 8AF  
Contact number: + 44 (0) 207 848 0323

Email addresses: [elizabeth.ryan@kcl.ac.uk](mailto:elizabeth.ryan@kcl.ac.uk) (Elizabeth G. Ryan), [c.drovandi@qut.edu.au](mailto:c.drovandi@qut.edu.au) (Christopher C. Drovandi), [a.pettitt@qut.edu.au](mailto:a.pettitt@qut.edu.au) (Anthony N. Pettitt)

*Keywords:* Bayesian optimal design; Nonlinear mixed effects models; Population design; Sampling strategies; Stochastic optimisation

---

## 1. Introduction

### 1.1. Background

Nonlinear mixed effects models (NLMEMs) are commonly used to model data in which heterogeneity exists between study subjects. For example, population pharmacokinetic (PK) studies investigate the disposition of a drug within a large sample of subjects. For an extensive overview of NLMEMs, as well as general theoretical developments and examples, see Racine-Poon (1985) and Davidian (2009). NLMEMs require the mean value of the population parameters to be estimated, as well as their inter-individual variability. The experimental design, which is usually under the control of the investigator, is responsible for determining the quality of the analyses of the data modelled by the NLMEM. These models, which are also known as the *population approach*, can allow for a sparse sampling design where only a few data points are available per individual, but a large number of individuals are included in the study. This is useful for studies in which the experimenters are interested in collecting informative data to obtain precise parameter estimates, but the number of samples per subject is limited due to ethical, physiological, time or cost constraints.

As the use of NLMEMs for modelling data from population studies has increased, so has the importance of optimally designing population studies so that accurate and precise estimates of the population parameters can be obtained (e.g., Mentré et al. (1997); Retout and Mentré (2003); Han and Chaloner (2004)). In this paper we are interested in (static) experimental designs that are optimal for the estimation of the population parameters in NLMEMs.

It is well known that the sampling times in a PK study can have a large impact on the precision and bias of parameter estimates (D’Argenio (1981)). However, there is contention in the literature as to whether it is better to sparsely sample a larger number of individuals, or to heavily sample a smaller number of individuals in population studies. Sheiner and Beal (1983), Hashimoto and Sheiner (1991), and Jonsson et al.

(1996) evaluate the effect of altering the number of subjects, and the number and timing of blood samples in PK studies on the precision and bias of the estimated parameter values. Sheiner and Beal (1983), and Hashimoto and Sheiner (1991) recommend that designs which use more study subjects, even if the majority are sparsely sampled, are preferable over designs which use more sampling times on fewer individuals. Conversely, Jonsson et al. (1996) recommend increasing the number of samples per subject, even if the total number of study subjects is small, so that the parameter estimates are unbiased and precise. However, these studies only investigate the use of a small number of sampling times (up to three sampling times), which were often fixed in their values and did not optimise a utility function over the design space.

### 1.2. Bayesian Hierarchical Model Framework

In the Bayesian framework, mixed effects models are commonly constructed using hierarchical models. These models account for the different levels of variability within and between populations. The observable outcomes are modelled conditionally on certain parameters, which are themselves assigned a probability distribution in terms of other parameters which are known as *hyperparameters*.

Here we consider NLMEMs where the  $h$ -th observation of individual  $i$ ,  $Y_{i,h}$ , is given by:

$$Y_{i,h} = f(\boldsymbol{\phi}_i, \mathbf{d}_{i,h}) + \epsilon_{i,h},$$

where  $f(\cdot)$  is a nonlinear mean function (which is the same for all individuals);  $\boldsymbol{\phi}_i$  is a random (or subject-specific) effect for individual  $i$ , which is treated as a latent variable in a Bayesian analysis;  $\mathbf{d}_{i,h}$  is the experimental setting; and the errors are independent and distributed as  $\epsilon_{i,h} \sim N(0, \sigma^2)$ . It is assumed that there are  $n$  subjects involved in the study.

A population distribution is specified for the subject-specific effects  $\boldsymbol{\phi}_i$ ,  $i = 1, \dots, n$ :

$$\boldsymbol{\phi}_i \sim MVN(\boldsymbol{\phi}, \boldsymbol{\Omega}),$$

where  $\boldsymbol{\phi}$  and  $\boldsymbol{\Omega}$  are the population mean and variance-covariance matrix respectively. Here  $\boldsymbol{\phi}$  is the population parameter or fixed effect. In this work we assume  $\boldsymbol{\Omega}$  is known. Racine-Poon (1985) provides guidance on prior elicitation for unknown  $\boldsymbol{\Omega}$ . We define

the unknown quantities as  $\theta = (\phi, \phi_{1:n})$ , where  $\phi_{1:n} = (\phi_1, \dots, \phi_n)$ . A log normal or inverse gamma prior (with shape hyperparameters  $a_0$  and  $b_0$ ) may be used for the observational variance,  $\sigma^2$ . Priors for the population parameters may also be specified as:

$$\phi \sim MVN(\mu, \Sigma).$$

where  $\mu$  is the prior mean for  $\phi$  and  $\Sigma$  is the prior variance-covariance for  $\phi$ . In this work, the values for the hyperparameters  $\mu$  and  $\Sigma$  are based on historical data.

### 1.3. Bayesian Optimal Design Theory

Bayesian experimental design has recently gained popularity in the literature and has many real-world applications. The Bayesian optimal design framework involves defining a prior distribution for the population parameter  $\phi$  and a population model for the subject-specific effects  $\phi_{1:n}$ ; a conditional sampling distribution  $p(\mathbf{y}_i|\mathbf{d}, \phi_i)$  for observing a new set of measurements  $\mathbf{y}_i$  for individual  $i$  at the design points  $\mathbf{d}$ , given parameter values  $\phi_i$ ; and a utility function  $U(\mathbf{d}, \theta, \mathbf{y})$  that describes the reward that is obtained for taking measurements at design points  $\mathbf{d}$ , and observing the data  $\mathbf{y}$  (where  $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ ), assuming the model parameter values  $\theta$  are known. In this work, we assume that measurements are taken at the same design points  $\mathbf{d}$  for all study subjects.

The optimal Bayesian design,  $\mathbf{d}^*$ , maximises the expected utility function  $U(\mathbf{d})$  over the design space  $\mathbf{D}$  with respect to the future data  $\mathbf{y}$  and model parameter  $\theta$ :

$$\begin{aligned} \mathbf{d}^* &= \arg \max_{\mathbf{d} \in \mathbf{D}} E\{U(\mathbf{d}, \theta, \mathbf{y})\} \\ &= \arg \max_{\mathbf{d} \in \mathbf{D}} \int_{\mathbf{Y}} \int_{\Phi} \int_{\Phi_{1:n}} U(\mathbf{d}, \theta, \mathbf{y}) \prod_{i=1}^n \left\{ p(\mathbf{y}_i|\mathbf{d}, \phi_i) p(\phi_i|\phi) d\phi_i \right\} p(\phi) d\phi d\mathbf{y}, \quad (1) \end{aligned}$$

where  $p(\mathbf{y}_i|\mathbf{d}, \phi_i) p(\phi_i|\phi)$  is the complete data likelihood for subject  $i$ ,  $p(\mathbf{y}_i|\mathbf{d}, \phi)$  is the observed data likelihood for subject  $i$ ,  $p(\phi_i|\phi)$  is the population distribution for the random effects and  $p(\phi)$  is the prior distribution for the fixed effects  $\phi$ . The random effects must be integrated out from equation (1).

The applied statistician's default approach to analysing data in a Bayesian framework is to use uninformative priors. When designing an experiment, it is reasonable to assume that, generally, there is historical and/or expert opinion available to formulate an informative prior (e.g., Clyde et al. (1996); Stroud et al. (2001); Ryan et al.

(2015)). Therefore, the prior predictive distribution gives ranges of data values that the experimenter believes to be plausible.

#### 1.4. Likelihood Function Approximation for NLMEMs

Due to the nonlinearity, NLMEMs have no analytical expression for the observed data likelihood. There is a wealth of literature on approaches for approximating the observed data likelihood to perform inference for the population parameters  $\phi$ . These methods include: first-order approximations (Sheiner and Beal (1983)); first order conditional methods (Lindstrom and Bates (1990)); Gaussian quadrature (e.g., Davidian and Gallant (1993)); adaptive Gaussian quadrature (e.g., Rabe-Hesketh et al. (2004)); Laplace approximations (e.g., Beal and Sheiner (2002)); Markov chain Monte Carlo (MCMC) (e.g., Spiegelhalter et al. (1996)); Monte Carlo integration (e.g., Wakefield (1994)); and importance sampling (Geweke (1989)). From the Bayesian perspective, the unknowns  $\phi_{1:n}$  and  $\phi$  have the same status, i.e., inference for either  $\phi_{1:n}$  or  $\phi$  is made using the posterior distribution. This is not the case for frequentist inference where  $\phi_{1:n}$  are random variables and  $\phi$  is a constant.

#### 1.5. Bayesian Designs for NLMEMs

Classical design criteria often consist of scalar functions of the Fisher information matrix (FIM), such as the determinant or the trace (e.g., Mentré et al. (1997); Retout and Mentré (2003)). Pseudo-Bayesian design criteria also rely on functions of the FIM, but also average these functions over a “prior” for the model parameters to account for parameter uncertainty (e.g., Pronzato and Walter (1985)). Once an expression for the likelihood has been found one can then derive the FIM.

Bayesian design criteria are often based upon the expected gain in Shannon information from the prior to posterior distribution (also known as ‘mutual information’ or the ‘Kullback-Leibler distance’) (e.g., Chaloner and Verdinelli (1995)). Other commonly-used Bayesian design criteria are based on the spread of the posterior distribution, which may be measured, for example, by the precision or by the entropy (e.g., Stroud et al. (2001)). When the posterior distribution is found by simulation, it must be sampled from for each future data set that is drawn from the prior predictive distribution, and so many thousands of posterior distributions are often required to perform

Bayesian experimental design. For this reason, fully Bayesian experimental designs for NLMEMs are largely unexplored.

Han and Chaloner (2004) searched for Bayesian population designs for a HIV dynamics study. They did not optimise over a continuous design space, but instead considered 8 fixed sampling schedules. The posterior predictive variance for two parameters was used in the utility function, and MCMC was used to sample the posterior distribution for each future dataset. Palmer and Müller (1998) implemented Bayesian optimal designs for population models for determining the timing of stem cell collections in cancer patients. The NLMEM parameters were estimated by MCMC simulation and a discrete set of designs was searched over.

Stroud et al. (2001) fitted NLMEMs to existing data relating to the PK of the anticancer agent, paclitaxel, in patients and found (sequential) Bayesian designs for the subject-specific parameters for the next patient. The design variable was the blood sampling times and two utility functions were used: the posterior precision of the area under the curve and the posterior precision of the time above a certain drug concentration. Each of these utility functions also included a cost penalty (which penalised sampling times that occurred after some pre-specified time) and were estimated using importance sampling. A Metropolis-Hastings MCMC algorithm (e.g., Müller (1999)) was used to find the optimal design.

### *1.6. Contribution and Outline*

In this paper, we present static, fully Bayesian designs for population parameters of NLMEMs, in which we use simulation-based optimal design methods to search over both continuous and discrete design spaces. We define a ‘continuous design space’ to be one in which the designs can take on any value in a pre-defined continuous interval, rather than values from a fixed set of discrete values. Whilst previous studies have found optimal Bayesian designs for NLMEMs by searching over a finite set of designs (e.g., Han and Chaloner (2004); Palmer and Müller (1998)), to our knowledge, no studies have searched over a continuous design space to find optimal static Bayesian designs for NLMEMs. This work is motivated by a PK study conducted by McGree et al. (2012) which models the PK of an acepromazine metabolite in racing horses. Our

design problem consists of finding the optimal sampling times for the PK study, as well as the optimal number of subjects to incorporate into the study, and the optimal number of samples to take per subject. To our knowledge, no previous Bayesian experimental design studies have addressed all of these issues (over a continuous design space). We are interested in finding designs that maximise the posterior precision of the population parameters, subject to cost constraints.

In Section 2 we describe the design methodology used in this work. In Section 3 we obtain some results for a simple example that involves designing for a linear mixed effects regression model. In Section 4 the PK case study is introduced and our design methods are applied to the case study in Section 5. The article concludes with a discussion in Section 6.

## 2. Bayesian Experimental Design Framework

Equation (1) does not usually have a closed form solution, and so numerical approximations or simulation methods are used to solve the maximisation and integration problem. These include: numerical quadrature or Laplace approximations (Brockwell and Kadane (2003); Ryan et al. (2015)); prior simulation (Müller (1999)); MCMC simulation in an augmented probability model (Müller (1999)); and sequential Monte Carlo (Amzal et al. (2006)).

In this work, we use the approach implemented by Müller (1999) to solve equation (1). This involves the use of MCMC which samples from the target distribution:

$$h(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y}) \propto U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d}),$$

using a Metropolis-Hastings scheme.  $h(\cdot)$  is constructed in such a way that the marginal distribution  $h(\mathbf{d})$  is proportional to the expected utility,  $U(\mathbf{d})$ . It is assumed that the utility  $U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$  is non-negative and bounded, and that  $h(\cdot)$  is integrable and can be normalised. The sample of simulated  $\mathbf{d}$  may be used to provide an estimate of  $h(\mathbf{d})$  and the joint mode of  $h(\mathbf{d})$ ,  $\mathbf{d}^*$ , corresponds to the optimal design. The Metropolis-Hastings scheme of Müller (1999) is described in Algorithm 1 and has been adapted for designing for NLMEMs.



---

**Algorithm 1:** MCMC algorithm for Bayesian optimal design for NLMEMs

---

- 1 Set an initial design  $\mathbf{d}^{(1)}$ .
- 2 Draw  $\phi \sim p(\phi)$ ,  $\phi_i \sim p(\phi_i|\phi)$ ,  $\mathbf{y}_i \sim p(\mathbf{y}_i|\mathbf{d}^{(1)}, \phi_i)$ , for  $i = 1, \dots, n$  individuals.
- 3 Compute  $U^{(1)} = U(\mathbf{d}^{(1)}, \theta, \mathbf{y})$ , where  $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$  and  $\theta = (\phi, \phi_{1:n})$ .
- 4 **for**  $j = 1$  **to**  $iters$  **do**
- 5     Generate a candidate design  $\tilde{\mathbf{d}}$  from a proposal distribution  $q(\cdot|\mathbf{d}^{(j)})$ .
- 6     Generate proposals for the parameters and simulate data:
- 7      $\tilde{\phi} \sim p(\phi)$ ,  $\tilde{\phi}_i \sim p(\phi_i|\tilde{\phi})$ ,  $\tilde{\mathbf{y}}_i \sim p(\mathbf{y}_i|\tilde{\mathbf{d}}, \tilde{\phi}_i)$ , for  $i = 1, \dots, n$  individuals.
- 8     Compute  $\tilde{U} = U(\tilde{\mathbf{d}}, \tilde{\theta}, \tilde{\mathbf{y}})$ , where  $\tilde{\mathbf{y}} = (\tilde{\mathbf{y}}_1, \dots, \tilde{\mathbf{y}}_n)$  and  $\tilde{\theta} = (\tilde{\phi}, \tilde{\phi}_{1:n})$ .
- 9     Calculate the MH acceptance probability,  $a = \min(1, A)$  where

$$A = \frac{\tilde{U} \times q(\mathbf{d}^{(j)}|\tilde{\mathbf{d}})}{U^{(j)} \times q(\tilde{\mathbf{d}}|\mathbf{d}^{(j)})}.$$

Here  $U^{(j)}$  and  $\mathbf{d}^{(j)}$  are the current utility and design point values, respectively, and  $\tilde{U}$  and  $\tilde{\mathbf{d}}$  are the proposed utility and design point values, respectively.

- 10     Set
 
$$(\mathbf{d}^{(j+1)}, U^{(j+1)}) = (\tilde{\mathbf{d}}, \tilde{U})$$

with probability  $a$ , and

$$(\mathbf{d}^{(j+1)}, U^{(j+1)}) = (\mathbf{d}^{(j)}, U^{(j)})$$

with probability  $1 - a$ .
  - 11 **endfor**
- 

Simulation-based algorithms such as those presented by Müller (1999) have been found to have slow convergence for situations where there are a large number ( $\geq 4$ ) of design variables (e.g., Stroud et al. (2001); Amzal et al. (2006)). For our PK application of interest, we are interested in searching for up to 15 sampling times. To ease the computational burden of having to search for a large number of design points, we use an approach from our previous work (Ryan et al. (2014)) which involves a lower dimensional parameterisation that reduces the design problem to one that involves searching over two design variables. The lower dimensional parameterisation was used in Line 5 in Algorithm 1. The sampling times for the PK study (Sections 4 and 5) will be generated from the evenly spaced percentiles of a  $Beta(a, b)$  distribution (see Ryan et al. (2014)), scaled to  $[0, 48]$  hours, where  $a, b > 0$ . Using this lower dimensional parameterisation, the Müller (1999) algorithm searches over the two design variables  $(a, b)$ , and once these optimal values are found, a large number of design points can be generated from the evenly-spaced percentiles of the  $Beta(a, b)$  distribution. However, it must be noted that the designs generated by this lower dimensional parameterisation

are not optimal but *near* optimal, which is a compromise of the computational savings achieved through these methods. We chose this lower dimensional parameterisation as we have used it previously for designing for fixed effects PK models (see Ryan et al. (2014)) and found that it gave fairly flexible designs that could be suitable for PK studies.

MCMC runs of 10000 iterations were performed and the convergence of the MCMC algorithm was carefully monitored (through examination of autocorrelation plots, histograms and contour plots of the design variables, and trace plots of the utility functions over the iterations). To determine the optimal designs, we searched for the multivariate mode of the multivariate normal kernel smoothing density estimates of the design variables (see Cook et al. (2008); Drovandi and Pettitt (2013)).

### 2.1. Utility Function Estimation via Importance Sampling

Utility functions are problem-specific and incorporate the aims of an experiment. Bayesian utility functions make recourse to the posterior distribution  $p(\theta|\mathbf{d}, \mathbf{y})$ . However, for NLMEMs, the posterior does not have a closed form expression and numerical methods are required for its approximation. We use a similar approach to Stroud et al. (2001) to generate samples from the posterior distribution, via importance sampling, for use in the Bayesian utilities.

Importance sampling is a commonly-used approach for approximating target distributions (Geweke (1989)) (in this case, the posterior  $p(\phi_{1:n}, \phi|\mathbf{d}, \mathbf{y})$ ). It involves choosing an *importance* distribution  $g(\cdot)$  from which it is easy to sample, and then weighting the samples to account for any differences between the importance and target distribution. The target and importance distributions should have the same support. Weighted discrete approximations,  $\{(\phi, \phi_{1:n})^{(l)}, W^{(l)}\}_{l=1}^{M_p}$ , are produced (where  $M_p$  is the number of values or particles used) from the target distribution, where

$$w(\phi, \phi_{1:n}) = \frac{\left\{ \prod_{i=1}^n p(\mathbf{y}_i|\mathbf{d}, \phi_i) p(\phi_i|\phi) \right\} p(\phi)}{g(\phi, \phi_{1:n})}$$

are the importance weights, and  $W^{(l)} \propto w((\phi, \phi_{1:n})^{(l)})$  are the normalised importance weights,  $\sum_{l=1}^{M_p} W^{(l)} = 1$ .

We use the population distribution (for the random effects) and the prior (for the fixed effects) as the importance distribution ( $g(\boldsymbol{\phi}, \boldsymbol{\phi}_{1:n}) = \left\{ \prod_{i=1}^n p(\boldsymbol{\phi}_i | \boldsymbol{\phi}) \right\} p(\boldsymbol{\phi})$ ), which reduces importance weights to the conditional distribution of the data given all the parameters:

$$w(\boldsymbol{\phi}, \boldsymbol{\phi}_{1:n}) = \prod_{i=1}^n p(\mathbf{y}_i | \mathbf{d}, \boldsymbol{\phi}_i).$$

This is not to be confused with the observed data likelihood  $p(\mathbf{y}_{1:n} | \mathbf{d}, \boldsymbol{\phi})$  in which the random effects have been integrated out. Note that the prior  $p(\boldsymbol{\phi})$  for the population parameters is relatively informative for our application as it is based on the results from the analysis of previous experiments.

To measure the efficiency of importance sampling, the effective sample size (ESS) is used (Geweke (1989)), where

$$ESS = \frac{1}{\sum W^{(l)2}}, 1 \leq ESS \leq M_p.$$

In this work we are only interested in the posterior distribution for the population parameters, and so we only use the samples of the subject-specific effects  $\boldsymbol{\phi}_{1:n}$  to calculate the importance weights and discard them thereafter. Our method for approximating the posterior distributions for the population parameters  $p(\boldsymbol{\phi} | \mathbf{d}, \mathbf{y})$  is outlined in Algorithm 2.

---

**Algorithm 2:** Algorithm for approximating  $p(\boldsymbol{\phi} | \mathbf{d}, \mathbf{y})$

---

- 1 Draw  $\{\boldsymbol{\phi}^{(l)}\}_{l=1}^{M_p}$  from  $p(\boldsymbol{\phi})$ .
  - 2 **for**  $i = 1 : n$  **do**
  - 3     Draw  $\{\boldsymbol{\phi}_i^{(l)}\}_{l=1}^{M_p}$  from  $p(\boldsymbol{\phi}_i | \boldsymbol{\phi}^{(l)})$ ,  $l = 1, \dots, M_p$
  - 4 **endfor**
  - 5  $\{\boldsymbol{\phi}^{(l)}\}_{l=1}^{M_p}$  and  $\{\boldsymbol{\phi}_{1:n}^{(l)}\}_{l=1}^{M_p}$  are only drawn once at the beginning of Algorithm 1 (prior to line 1) and are stored.
  - 6 Weight  $w^{(l)} = \prod_{i=1}^n p(\tilde{\mathbf{y}}_i | \mathbf{d}, \boldsymbol{\phi}_i^{(l)})$ ,  $l = 1, \dots, M_p$ .
  - 7 Normalise  $w^{(l)}$  to give  $W^{(l)}$ ,  $l = 1, \dots, M_p$ .
  - 8 The particle approximation to  $p(\boldsymbol{\phi} | \mathbf{d}, \mathbf{y})$  is given by  $\{\boldsymbol{\phi}^{(l)}, W^{(l)}\}_{l=1}^{M_p}$ .
- 

The utility functions can then be estimated using the weighted samples. Algorithm 2 (lines 6 - 8) is used in each iteration of Algorithm 1 (line 8, as well as line 3 at the beginning of the algorithm) to calculate the utility function. For our applications, we

will use the determinant of the posterior precision matrix of the population parameters as the utility function:

$$U(\mathbf{d}, \mathbf{y}) = \det(\text{prec}(\boldsymbol{\phi}|\mathbf{d}, \mathbf{y})).$$

For the example considered in Section 5, we set  $M_p = 100000$  as this number of particles provided reasonably stable (based on the ESS) and precise estimates of the utility function.

### 3. Simple Illustrative Example: Linear Mixed Effects Model

We will begin with a simple example, in which we determine the optimal balance between the number of subjects ( $n$ ) and the number of samples per subject ( $n_d$ ), as well as the optimal values for the predictor variable  $\mathbf{x}$ , under certain conditions. The model is a linear mixed effects model where the response for the  $i$ -th subject is given by:

$$\mathbf{y}_i = \mathbf{X}\boldsymbol{\phi} + \mathbf{Z}_i\boldsymbol{\phi}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n.$$

The observation vector  $\mathbf{y}$  is of dimension  $N \times 1$  (where  $N = n \times n_d$ , i.e., the total number of observations).  $\mathbf{X}$  is the design matrix for the fixed effects (of dimension  $n_d \times p$ ) and  $\mathbf{Z}_i$  is the design matrix for the  $i$ -th subject (of dimension  $n_d \times p$ ).  $\boldsymbol{\phi}$  consists of the  $(p \times 1)$  fixed effects, with the prior  $\boldsymbol{\phi} \sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$ , where  $\boldsymbol{\Sigma}$  is a known,  $p \times p$  nonsingular matrix.  $\boldsymbol{\phi}_{1:n}$  consists of the  $(np \times 1)$  random effects, with the model:

$$(\boldsymbol{\phi}_1^T, \dots, \boldsymbol{\phi}_n^T) \sim MVN(\mathbf{0}, \boldsymbol{\Omega}),$$

where  $\boldsymbol{\Omega}$  is a known,  $np \times np$  nonsingular matrix. It is assumed that the random effects are independent. The residuals,  $\boldsymbol{\epsilon}_i$ , are independently distributed with  $\boldsymbol{\epsilon}_i \sim MVN(\mathbf{0}, \mathbf{I}\sigma^2)$ ,  $i = 1, \dots, n$ . Here we assume that  $\sigma^2$  is known. It is also assumed that the observational errors  $\boldsymbol{\epsilon}_i$  are independent of the fixed or random effects.

For the linear mixed effects model, Sorenson and Gianola (2002) have derived an analytical expression for the posterior density of the fixed effects and we will use their expression (see later, equation (4)) for the posterior precision matrix to construct our Bayesian utility function. Since the integrals in (1) can be computed analytically, all that is required is to perform the optimisation to find the design which maximises the utility function.

We use a linear regression model of the following form:

$$y_{i,h} = (\phi_0 + \phi_{i,0}) + (\phi_1 + \phi_{i,1})x_{i,h} + (\phi_2 + \phi_{i,2})x_{i,h}^2 + \epsilon_{i,h},$$

where  $i = 1, \dots, n$ ,  $h = 1, \dots, n_d$ .  $\boldsymbol{\phi}^T = (\phi_0, \phi_1, \phi_2)$  and  $\boldsymbol{\phi}_i^T = (\phi_{i,0}, \phi_{i,1}, \phi_{i,2})$ .

We set  $\boldsymbol{\Sigma} = \text{diag}(0.4, 0.3, 0.2)$ ;  $\boldsymbol{\Omega} = \text{diag}(0.2, 0.15, 0.12, 0.2, 0.15, 0.12, \dots, 0.2, 0.15, 0.12)$ , where the number of repeats of the variances (0.2, 0.15, 0.12) for the three random effects depends on the number of subjects,  $n$ ; and  $\sigma^2 = 0.01$ .

The utility function is the (log) determinant of the posterior precision of the population parameters:

$$U(\mathbf{d}, \mathbf{y}) = \log(\det(\text{prec}(\boldsymbol{\phi}|\mathbf{d}, \mathbf{y}))). \quad (2)$$

In addition to precisely estimating the model parameters, our design objectives also included a cost constraint. Since the posterior precision (of the population parameters) and cost penalty may not be on the same scale, equation (2) was used to find designs, subject to a certain fixed maximum cost (Stigler (1971)).

The cost function penalised for the number of subjects in the study, and the total number of measurements taken in the study:

$$C = c_{sub} \cdot n + c_{sample} \cdot n \cdot n_d, \quad (3)$$

where  $c_{sub}$  is the cost per subject, and was set to \$50, and  $c_{sample}$  is the cost per sample, and was set to \$10. These values were arbitrarily chosen for illustrative purposes. The cost function was used to determine the different combinations of the number of subjects and samples per subject that could be included in the study for a fixed total cost. The (log) determinant of the posterior precision was calculated for each of these combinations and comparisons were made to see which yielded the highest value of the utility function.

The expression for the posterior precision of  $\boldsymbol{\phi}$  is given by:

$$\text{prec}(\boldsymbol{\phi}|\mathbf{d}, \mathbf{y}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} + \boldsymbol{\Sigma}^{-1}\sigma^2)\sigma^{-2}, \quad (4)$$

where  $\mathbf{V} = \mathbf{Z}\boldsymbol{\Omega}\mathbf{Z}'\sigma^{-2} + \mathbf{I}$  (Sorenson and Gianola (2002)). Here,  $\mathbf{Z}$  denotes the design matrix for the random effects (of dimension  $N \times np$ ). If  $\boldsymbol{\Omega} = \mathbf{0}$ , then  $\mathbf{V} = \mathbf{I}$ . Note

that, although the posterior precision does not depend on  $\mathbf{y}$  here, the posterior mean is dependent on  $\mathbf{y}$ . The designs  $\mathbf{d}$  enter the utility function via the design matrices  $\mathbf{X}$  and  $\mathbf{Z}$ , and therefore only the term  $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$  in the right hand side of equation (4).

### 3.1. Results

For a fixed cost of \$750, we investigated different combinations of the number of subjects and the number of samples per subject, to determine the optimal balance between these two quantities. We also searched over the design space for the predictor  $\mathbf{x}$ , whose values were restricted to occur between 0 and 1.

The optimal exact designs were found using the Müller (1999) algorithm with 10000 iterations. According to classical design theory, the D-optimal design for a quadratic equation with three unknown fixed effects parameters should have three support points - one at either end of the design space and one at the centre (e.g., Pukelsheim (1993); Tan and Berger (1999)). Therefore, we decided to search over three predictor design variables,  $(x_1, x_2, x_3)$ , which were the three support points, and two weights/integers for an exact design  $(\omega_1, \omega_2)$ , where  $\omega_3 = n_d - (\omega_1 + \omega_2)$ . The weights determine the number of replicates that are to be placed on each support point. Therefore,  $\mathbf{d} = (x_1, x_2, x_3, \omega_1, \omega_2)$ . We also searched over 4 predictor design variables, but found that one of these design points was a replicate of one of the three support points.

Since an analytical expression of the utility function,  $U(\mathbf{d})$ , was available, the mode could easily be estimated by choosing the sample with the highest  $U(\mathbf{d})$  value. To simplify matters, it was assumed that all subjects had the same number of observations at the same values of the predictor variable  $\mathbf{x}$ . The results are summarised in Table 1.

From Table 1 it can be seen that the optimal support points for the model occur at the middle of the design space (0.5), and at either end (0 and 1). This is similar to results obtained in the classical design literature (e.g., Pukelsheim (1993)). Preference for the replicates was given to the centre (0.5) of the design space, followed by the start of the design space (0).

It can also be determined from Table 1 that, for this application of interest, it is preferable to take a smaller number of samples from a larger number of individuals

No. Sub- jects	No. samples per subject	Total no. samples	Optimal exact design for $x^\dagger, \xi$	Utility function value $U(d)$
9	3	27	$\xi = \begin{Bmatrix} 0 & 0.5 & 1 \\ 1 & 1 & 1 \end{Bmatrix}$	15.55
8	4	32	$\xi = \begin{Bmatrix} 0 & 0.5 & 1 \\ 1 & 2 & 1 \end{Bmatrix}$	15.47
7	5	35	$\xi = \begin{Bmatrix} 0 & 0.5 & 1 \\ 2 & 2 & 1 \end{Bmatrix}$	15.26
6	7	42	$\xi = \begin{Bmatrix} 0 & 0.5 & 1 \\ 2 & 3 & 2 \end{Bmatrix}$	14.77
5	10	50	$\xi = \begin{Bmatrix} 0 & 0.5 & 1 \\ 3 & 4 & 3 \end{Bmatrix}$	14.72
4	13	52	$\xi = \begin{Bmatrix} 0 & 0.5 & 1 \\ 4 & 6 & 3 \end{Bmatrix}$	14.27

Table 1: Optimal designs and utility function values for different combinations of the number of subjects and the number of samples per subject for a fixed cost of \$750. <sup>†</sup>Note that the weights for the exact design (row 2 of  $\xi$ ) correspond to the number of replicates that are to be taken on each support point (row 1 of  $\xi$ ) for 1 study subject. The sum of the weights is equal to the number of samples that are to be taken per subject.

(rather than heavily sample a smaller number of individuals). This is in agreement with Diggle et al. (1994) who note that for a uniform correlation structure (of the errors), the addition of one repeated measure within a subject conveys less information on the fixed effects than the addition of an independent measure of a new study subject.

#### *Prior Sensitivity*

We are now interested in investigating the effect of the prior distribution for the fixed effects and the model for the random effects on the optimal number of subjects and samples per subject. We will begin by varying the prior for the fixed effects  $\phi \sim N(0, c\Sigma)$ , where we will use the values  $c = 0.1, 1, 2, 10, 100$ . As  $c \rightarrow \infty$ , we would obtain the same results as in the frequentist paradigm. The population model for the random effects is the same as above. The same combinations of the number of subjects and samples per subject were used as in Table 1, and we will use the optimal  $\mathbf{x}$  values from this table.

From Figure 1, it can be seen that there does not appear to be any variation in the optimal number of samples to take per subject as the prior variance for the fixed effects changes. That is, the designs do not change with the prior variance for the fixed effects, but the values of the utility function decrease as the prior variance increases. It appears that it is more useful to take a smaller number of samples from a larger number of individuals to precisely estimate the fixed effects, regardless of how precise our  $a$

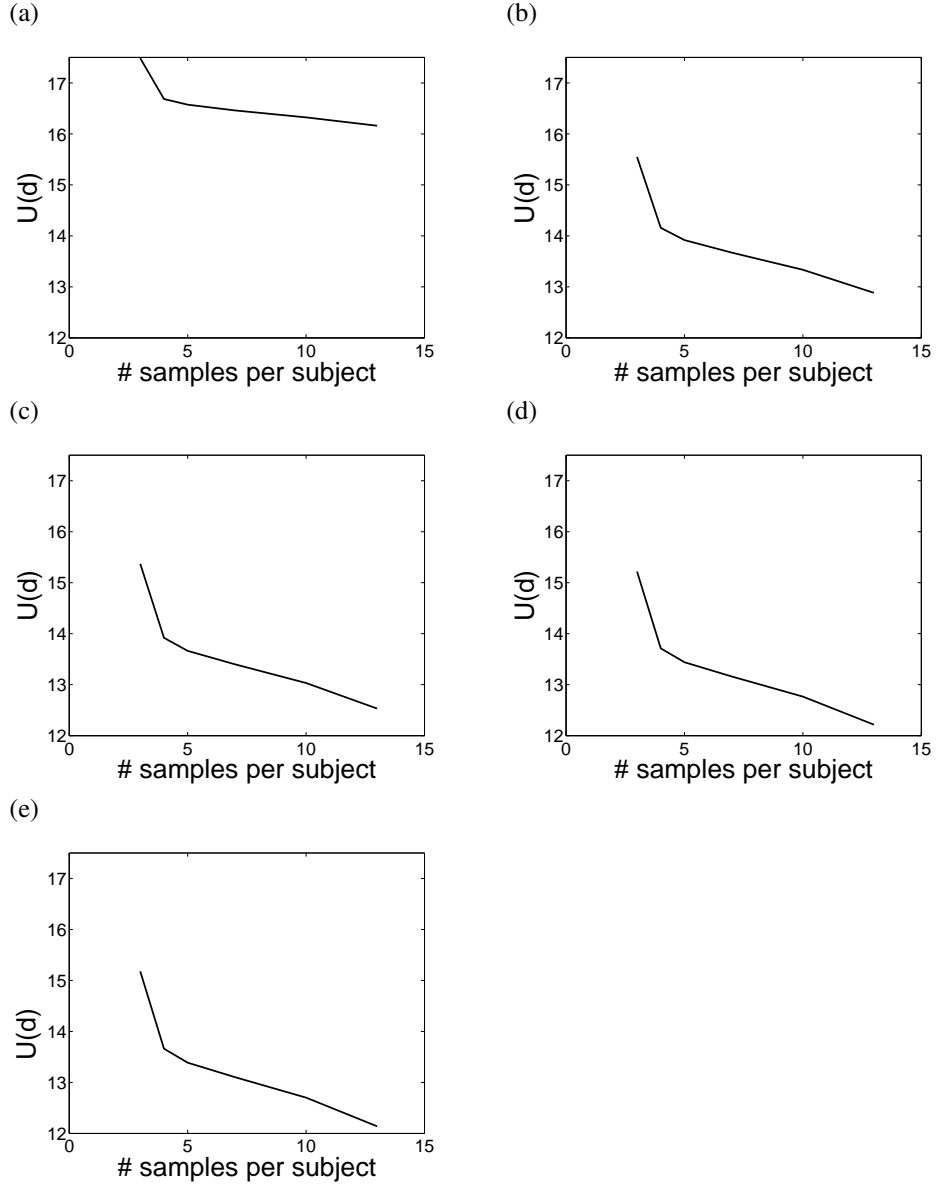


Figure 1: Utility function value versus the number of samples per subject, for a fixed cost of \$750, for various values of the prior variance for the population parameters for: (a)  $c = 0.1$ , (b)  $c = 1$ , (c)  $c = 2$ , (d)  $c = 10$ , and (e)  $c = 100$ .



*priori* knowledge of these effects is, for a fixed (and somewhat precise) value for the population variance of the subject-specific effects. This makes sense: since we already have precise knowledge about the subjects and wish to precisely estimate the population parameters, then we should incorporate more subjects into the study (even if they only have a few samples taken), regardless of our level of knowledge of the population parameters. Also, for a precise amount of knowledge of the population parameters ( $c = 0.1$ ) there is less variation in the utility values for the different combinations of the number of samples per subject and number of subjects.

It is important to remember, with respect to our utility function (equation (4)), that

$$\det(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} + \Sigma^{-1}\sigma^2) \neq \det(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}) + \det(\Sigma^{-1}\sigma^2).$$

This means that varying the prior variance of the fixed effects ( $c\Sigma$ ) may slightly affect the design when the determinant of the posterior precision is taken, where a smaller value of  $c$  will have a larger impact on the design compared to a larger value of  $c$ . If the trace of the posterior precision matrix was taken instead of the determinant, then varying the prior of the fixed effects would have no impact on the design since the trace is a linear function.

Now we investigate how changing the population variance of the subject-specific effects alters the designs:  $\phi_i \sim N(0, k\Sigma)$ , where  $k = 0.005, 0.05, 0.5, 1, 2, 20$ . For the population parameter, we will set the value of  $c$  to 10000, to give prior variance  $10000\Sigma$ , which essentially gives a frequentist analysis. If  $k \approx 0$ , then  $\text{Var}(\phi_i) \approx 0$ , and we will obtain the population parameter frequentist result. We will investigate the same combinations of the number of subjects and samples per subject as above, and use the optimal  $\mathbf{x}$  values from Table 1.

From Figure 2, it can be seen that as the population variance of the subject-specific effects increases, the optimal number of samples per subject decreases. That is, when there is a large amount of variation in the values of the subject-specific effects, one should focus on taking a small number of samples from a larger number of individuals, so that precise estimates of the population mean parameters can be obtained. If there is little variation in the subject-specific effects, then one should focus their resources on taking a larger number of samples from a smaller number of subjects, as there is little

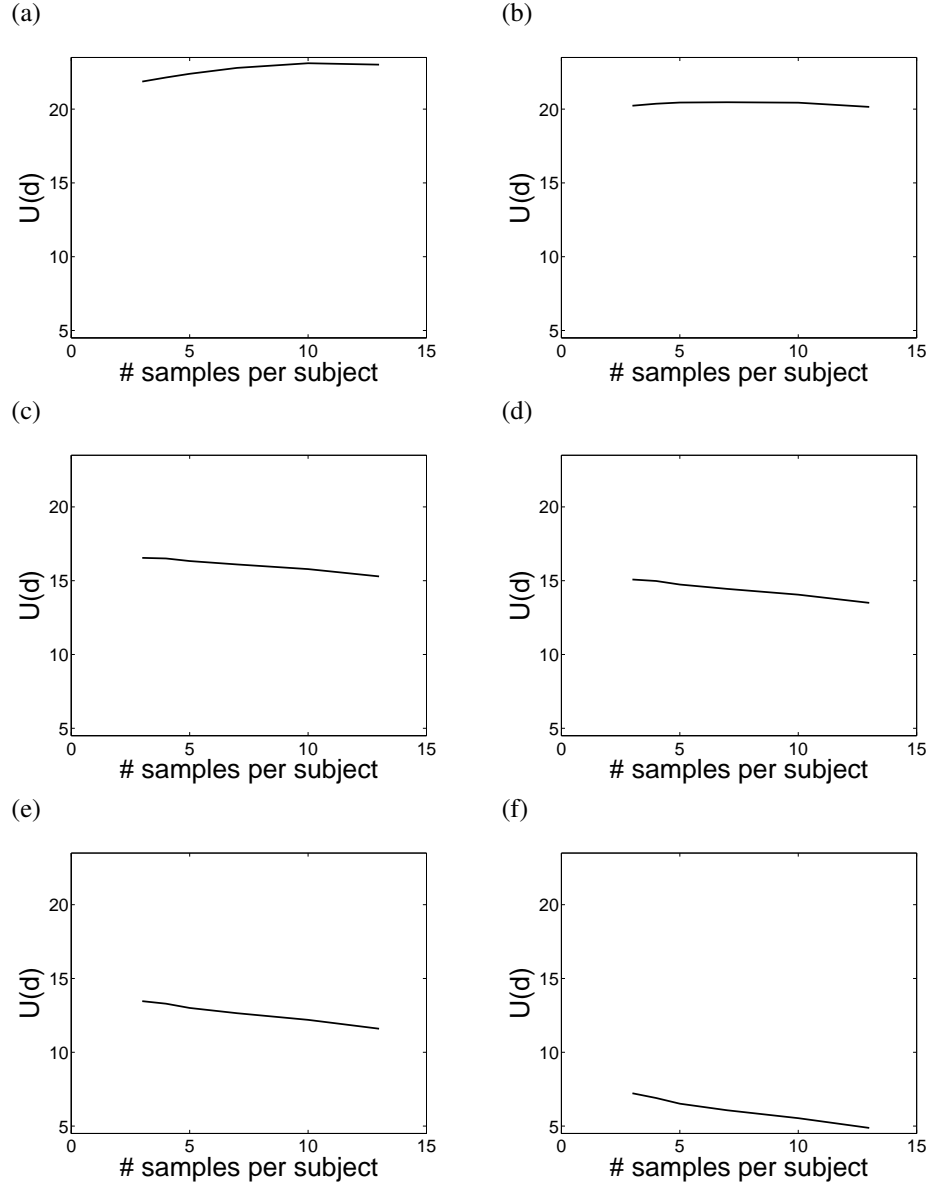


Figure 2: Utility function value versus the number of samples per subject, for a fixed cost of \$750, for various values of the variance for the subject-specific parameters: (a)  $k = 0.005$ , (b)  $k = 0.05$ , (c)  $k = 0.5$ , (d)  $k = 1$ , (e)  $k = 2$ , and (f)  $k = 20$ .

benefit from taking samples from more subjects to precisely estimate the population means.

An investigation into the sensitivity of the optimal design to the values chosen for the cost per subject and cost per sample was conducted and is presented in Online Appendix A. The results from the cost sensitivity analyses supported those of Table 1, in that it is preferable to take a small number of samples from a larger number of subjects, for this linear mixed effects model and priors.

It is important to note that, although the same fixed total cost is obtained for the different combinations in Table 1, the total number of observations taken is not the same for each combination. In Online Appendix B, we assume that the same total number of observations (here we assume 48) is taken (and that the costs per subject and sample are equivalent), and investigate the ‘best way’ to divide up these observations. It was found that the best way to divide up 48 observations was to take 3 samples each from 16 individuals, so that we could precisely estimate 3 population parameters.

#### **4. Case Study: Population Pharmacokinetics of HEPS in Horses**

This case study is concerned with determining the optimal urine sampling times for a population PK study of the acepromazine (ACP) metabolite 2-(1-hydroxyethyl) promazine (HEPS) in racing horses. We are also interested in determining the optimal number of horses to include in the study, and the optimal number of samples to take per horse (subject to cost constraints). Our case study will be a retrospective study design which makes use of the data collected and analysed by McGree et al. (2012). We are interested in re-designing the study to precisely estimate the mean PK parameters of the population of horses.

In the study conducted by McGree et al. (2012), 30mg of ACP was administered to 12 horses (geldings) and urine samples were taken at the following times: 2, 4, 6, 8, 12, 24, 36 and 48 hours after administration. The horses were trained to urinate to the sound of a whistle. Here, we will re-design the urine sampling times from those used in McGree et al. (2012) so that accurate measures of PK parameters of interest can be obtained.

#### 4.1. The model

We assume  $n_d$  urine samples,  $t_1, t_2, \dots, t_{n_d}$ , will be collected for  $n$  subjects. The cumulative amount of HEPS in subject  $i$ 's urine at the  $h$ -th sampling time,  $Y_{i,h}$ , is modelled by:

$$Y_{i,h} = f(\phi_i, t_h) + \epsilon_{i,h}, \quad i = 1, \dots, n; h = 1, \dots, n_d,$$

where

$$f(\phi_i, t_h) = \frac{D \times F_{E_i}(Cl_i - V_i \times k_a - Cl_i \times e^{-k_a t_h} + V_i \times k_a \times e^{-\frac{Cl_i}{V_i} t_h})}{Cl_i - V_i \times k_a},$$

and independent  $\epsilon_{i,h} \sim N(0, \sigma_{add}^2)$ .

Here  $\phi_i = (\log Cl_i, \log F_{E_i}, \log V_i)$  are the PK parameters for the  $i$ -th horse.  $Cl$  is the clearance rate,  $F_E$  is the fraction of HEPS that is excreted renally, and  $V$  is the volume of distribution (a theoretical measure). In this model it is assumed that  $k_a = 35.87$  is a constant (as per McGree et al. (2012)) and  $D = 30000\mu\text{g}$  is the drug dose. This model assumes a first-order absorption and elimination of the drug which is administered orally. The cumulative amount of the drug in the urine increases over time until all of the drug is eliminated.

Only additive error, whose variance is given by  $\sigma_{add}^2$ , is present in the model. It was assumed that  $\sigma_{add}^2 \sim N(1.2 \times 10^4, 3.1 \times 10^6)$  (based on McGree et al.'s (2012) results).

The priors were obtained from the posterior results of McGree et al. (2012), in which the Bayesian model above was fitted to the data using MCMC. The population distribution for the individual parameters  $\phi_i$ ,  $i = 1, \dots, n$ , is specified as:

$$\phi_i \sim MVN(\phi, \Omega),$$

where  $\phi = (\log Cl, \log F_E, \log V)$  is the population mean parameter and  $\Omega$  is the population variance-covariance matrix.  $\Omega$  is assumed to be known and was obtained from the results of McGree et al. (2012) and is given by:

$$\Omega = \begin{pmatrix} 0.0149 & 0.0034 & -0.0037 \\ 0.0034 & 0.0146 & -0.0027 \\ -0.0037 & -0.0027 & 0.0048 \end{pmatrix}.$$

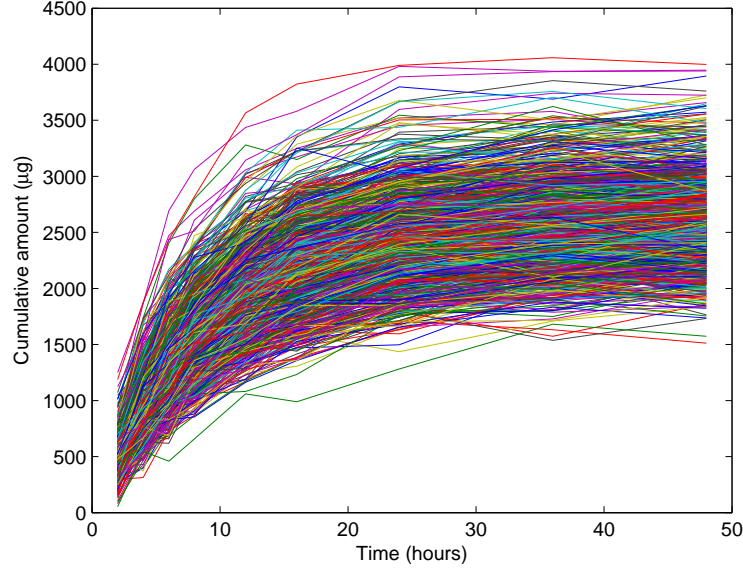


Figure 3: Prior predictive curves of cumulative urine amount of HEPS (online version in colour).

The fixed effects are assumed to have a prior distribution  $\phi \sim MVN(\mu, \Sigma)$ , with known mean and variance-covariance matrix:

$$\mu = \begin{pmatrix} 6.65 \\ -2.46 \\ 8.84 \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} 0.0076 & -0.0030 & 0.0050 \\ -0.0030 & 0.0050 & -0.0030 \\ 0.0050 & -0.0030 & 0.0073 \end{pmatrix}.$$

The values for  $\Omega$  and  $\Sigma$  were obtained from the posterior samples which fit the data for 12 horses and so are very precise.

The estimated posterior densities of  $\phi = (\log Cl, \log F_E, \log V)$  that were obtained by McGree et al. (2012) are displayed in Online Appendix C. These posteriors were used as our prior for the retrospective design by fitting a multivariate normal distribution to the MCMC output of McGree et al. (2012). Simulations from the prior predictive distribution for the cumulative urine amounts of HEPS are displayed in Figure 3.

No. Subjects	No. samples per subject	Total no. of samples	Fixed cost (\$)
2	10	20	\$3000
3	5	15	
4	2	8	
3	11	33	\$5000
4	7	28	
5	5	25	
6	3	18	
7	2	14	\$10 000
5	15	75	
8	7	56	
12	3	36	

Table 2: Different combinations of the number of subjects and the number of samples per subject for various fixed total costs that were explored for the PK example.

## 5. Designs for Population PK Study

We now turn our attention to determining the optimal tradeoff between the number of subjects and samples per subject, and the optimal sampling times for a PK study. The methods discussed in Section 2 will be applied to the horse PK study introduced in Section 4. Table 2 displays the combinations of the number of subjects and samples per subject that are investigated for fixed total costs of \$3000, \$5000, and \$10000.

The utility function is the determinant of the posterior precision of the population parameters and will be estimated using the procedure described in Section 2.1. We will implicitly use equation (3) as the cost function, where  $c_{sub}$  is the cost per horse, and was set to \$500, and  $c_{sample}$  is the cost per sample, and was set to \$100. These values were determined after consulting several horse PK experts. Separate MCMCs were run for each of the combinations of the number of horses and number of samples per horse (i.e., separate MCMCs were run in parallel for each line of Table 2).

### 5.1. Results

We search for the (near) optimal urine sampling times, which are restricted to occur between 0 and 48 hours and use the lower dimensional parameterisation discussed in Section 2. It was assumed that all of the horses were sampled at the same times following the administration of the drug. The results of the sampling times that were generated from the (evenly-spaced) percentiles of a  $Beta(a, b)$  distribution for the different combinations of the number of subjects and samples per subject are displayed

in Figure 4. The MCMC convergence diagnostics discussed in Section 2 were satisfied for all simulations that were performed, and Online Appendix D displays the convergence diagnostics for the MCMC simulations that were performed when  $n = 3$  and  $n_d = 11$  (i.e., the “3 horses, 11 sampling times per horse” combination). The most computationally intensive component of the MCMC algorithm was the estimation of the posterior distribution for use in the Bayesian utility function (i.e., Algorithm 2) since the posterior must be sampled from for each future dataset that is drawn from the prior predictive distribution. For this example, 100000 samples were required so that reasonably stable and precise estimates of the utility function could be obtained via importance sampling.

For each of the total fixed costs considered, it was found that it was preferable to heavily sample a smaller number of subjects (see Figure 4). That is, if one is interested in obtaining precise posterior distributions of the population (urine) PK parameters, then one should heavily sample (10-15 samples per individual) a small number, say 2-5, of individuals. From Figure 4, it can be seen that the more subjects included and samples taken in the study the better, as expected, provided there is no upper limit to the cost of the study (which would be rare in practice). However, these designs may not be suitable in practice as it may be difficult or unethical to take a large number, say, 15 urine samples from an individual within 48 hours. It should be noted that these conclusions are subject to the cost ratios used in this study.

Our results are in agreement with Jonsson et al. (1996), but are in contrast to the results obtained by Sheiner and Beal (1983), and Hashimoto and Sheiner (1991), who recommend that designs that use a larger number of subjects that are sparsely sampled are preferable over designs that use a smaller number of individuals that are heavily sampled. However, their designs are not fully Bayesian and have different design objectives to those considered in this paper. Also, these studies only consider up to three sampling times, which are fixed in their values, and do not optimise the sampling times over the design space as we have done here. These results highlight the fact that the optimal sampling strategy may not be obvious and optimisation of the design problem is required using the methods we have described.

The (near) optimal sampling times were evenly spread across the design space (see

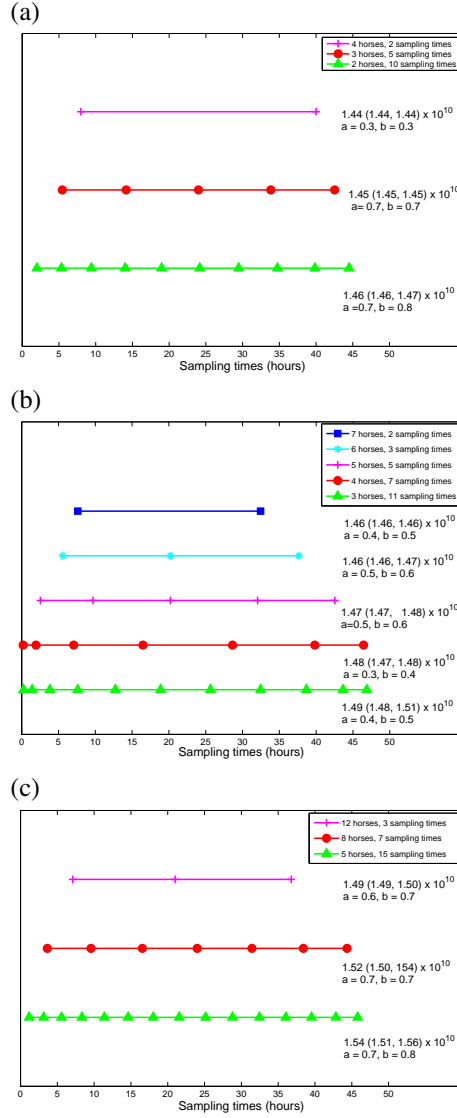


Figure 4: PK sampling times generated by the evenly spaced percentiles of a  $Beta(a, b)$  distribution, for the various combinations of the number of horses and number of sampling times per horse, for a fixed cost of (a) \$3000, (b) \$5000, and (c) \$10000. The utility function values are displayed next to the sampling times, along with the values for the shape parameters for the beta distribution that was used to generate the sampling times.



Figure 4). For a small number of sampling times, preference was given to sampling times that spanned the central region of the design space (e.g., 10 - 40 hours, Figure 4). This is the region of the PK curve where the increase in the cumulative amount of HEPS in urine begins to progress at a slower rate and eventually asymptotes (see Figure 3). As the number of sampling times increased, they were evenly spread out from the central region of the design space to cover a greater region of the PK curve. The earliest and latest sampling times occurred at 0.12 and 47.4 hours respectively, and were associated with the 5 horses, 15 sampling times combination. These sampling times covered the majority of the design space [0, 48] hours. For the majority of the population designs (where more than 3 samples were taken), sampling continued after the cumulative amount of drug had reached an asymptote.

It should be noted that there was little difference in the utility function values for the different combinations of the number of subjects and samples per subject, for a fixed total cost. In Figure 4(a) the utility function values ranged from 1.44 to 1.46, in Figure 4(b) they ranged from 1.46 to 1.49, and ranged from 1.49 to 1.54 in Figure 4(c) (on average). Therefore it is difficult to make strong statements on the optimal sampling strategies, based on these results. Alternative priors for the fixed effects or models for the random effects may produce utility surfaces that are less flat.

We also tried implementing Algorithm 1 to search for the optimal number of subjects, number of samples per subject, and sampling times all at once, but this was found to be too computationally intensive.

A prior sensitivity study was also conducted (in a similar fashion to Section 3.1) and can be found in Online Appendix E. It was found that the designs did not vary when the prior variance for the population parameters was altered (similar to Figure 1), and it was most useful to heavily sample a smaller number of subjects. When the variation between the subjects (population variance) was small, it was preferable to take more samples from a smaller number of individuals, as there was little benefit from sampling a larger number of individuals. When the variation between the subjects was larger, it was preferable to take a smaller number of samples from more subjects, so that precise posterior distributions of the population parameters can be obtained. This is similar to Figure 2.

## 6. Discussion

In this article we have discussed and presented methods that can be used to find optimal fully Bayesian designs for mixed effects models, for both linear and non-linear models. The design problem was to determine the optimal number of subjects, the optimal number of samples per subject and the optimal predictor variable values (as in the linear model example) or the near optimal sampling times for a PK study (nonlinear example), to precisely estimate the model parameter of interest, subject to a cost constraint. Whilst the computational methods used in this work are not novel, their adaptation and application to find fully Bayesian static optimal designs for NLMEMs are new. Searches over a number of different design variables (some of which had a continuous design space) were performed, which also has not been previously implemented to find fully Bayesian static designs for NLMEMs.

Population designs comprise of a set of elementary designs that are to be carried out on groups of subjects. The elementary designs consist of several values of the design variable (e.g., blood sampling times, treatment doses etc) that are to be performed on each subject belonging to the design. The number of samples to be taken and the values of the design variable may differ between subjects within an elementary design, and between the elementary designs. For simplicity, we assumed that all individuals in the examples considered in this paper had the same number of measurements taken and were sampled at the same experimental design. Previous simulation population PK studies (e.g., Sheiner and Beal (1983); Jonsson et al. (1996)) have found that the precision and accuracy of the parameter estimates are affected by the number of elementary designs, the number of subjects per elementary design, and the number and allocation of the design points (e.g., sampling times). Therefore, our design set up (in terms of having one elementary design) may not be optimal for population studies and future studies may wish to investigate the use of different elementary designs for different groups of subjects. This is likely to be very computationally intensive as many design variables would be involved.

In both the linear and nonlinear examples, we were also interested in determining the optimal number of subjects and samples per subject, subject to a cost constraint.

This was achieved by searching over several different combinations of the number of subjects and samples per subject that resulted in the same total fixed cost. We had adapted the Müller (1999) algorithm to treat the number of subjects and the number of samples per subject as design variables, so that the optimal number of subjects, samples per subject and sampling times could be found simultaneously. However, one cannot search over a large number of design variables using the Müller (1999) algorithm as it becomes too computationally intensive to search over the joint space  $(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$  and to determine the multivariate mode for the large number of design variables.

The results obtained in this study are dependent on the cost constraints and prior distributions used for the design problems, and for both examples, it is likely that different sampling strategies would have been achieved if different cost constraints or prior distributions were used. For the linear mixed effects model that was considered in this work, it was found that it was more useful to take a small number of samples from a larger number of individuals if one is interested in obtaining precise posterior distributions of the population parameters. For the NLMEM example (PK study), it was found that it was preferable to heavily sample a smaller number of individuals, so that precise posterior distributions of the population PK parameters could be obtained. The differences in the optimal sampling strategies between the two examples considered highlights how problem-dependent optimal Bayesian designs for mixed effects models are, emphasising the need for the optimisation methods presented in this paper.

A lower dimensional parameterisation was used to reduce the computational burden of searching over a large number of design points. The MCMC algorithm searched over the two design variables  $(a, b)$ , and once these optimal values are found, the design points were generated from the evenly-spaced percentiles of the  $Beta(a, b)$  distribution, scaled to  $[0, 48]$  hours. We have previously found the beta proposal scheme to be quite flexible in generating designs, in that a wide variety of designs can be generated from this scheme depending on the values of the shape parameters used (Ryan et al. (2014)). This parameterisation could be extended to offer further flexibility by including another design variable that determines the optimal percentiles of the beta distribution to use. It must be stressed that the designs generated by these lower dimensional parameterisations are not optimal but near optimal.

In the examples considered in this paper, we were able to estimate the posterior density (for use in the Bayesian utility functions) analytically, or via importance sampling. Future studies that design for mixed effects models in a Bayesian framework should investigate alternative methods for estimating the posterior density, such as adaptive Gaussian quadrature (e.g., Rabe-Hesketh et al. (2004)) or Laplace approximations (e.g., Wolfinger (1993)), which may prove to be computationally faster and more efficient. We found that importance sampling from the prior (fixed effects) and population distribution (random effects) was somewhat computationally intensive as many importance samples ( $M_p = 100000$ ) were required to obtain reasonably stable (based on the ESS) and precise estimates of the utility function.

### **Supplementary Materials**

Online Appendices that are referenced in Sections 3, 4 and 5 are available with this paper at the CSDA website on the ScienceDirect Online Library.

### **Acknowledgments**

E.G. Ryan was supported by an APA(I) Scholarship which came from an ARC Linkage Grant with Roche Palo Alto (LP0991602). The work of A.N. Pettitt was also supported by an ARC Discovery Project (DP110100159). The authors would like to thank Dr James McGree from Queensland University of Technology for his helpful comments throughout the process of writing this paper; for supplying the posterior samples for the horse PK model; and for consultation about his paper (McGree et al. (2012)). The authors would like to thank the reviewers and the editor for their helpful comments during the revision process of this manuscript.

### **References**

Amzal, B., Bois, F., Parent, E., Robert, C. P., 2006. Bayesian-optimal design via interacting particle systems. *Journal of the American Statistical Association* 101 (474), 773–785.

- Beal, S., Sheiner, L., 2002. NONMEM user's guide. Tech. rep., University of California, San Francisco, NONMEM Project group.
- Brockwell, A. E., Kadane, J. B., 2003. A gridding method for Bayesian sequential decision problems. *Journal of Computational and Graphical Statistics* 12 (3), 566–584.
- Chaloner, K., Verdinelli, I., 1995. Bayesian experimental design: A review. *Statistical Science* 10, 273–304.
- Clyde, M. A., Müller, P., Parmigiani, G., 1996. Exploring expected utility surfaces by Markov Chains. Tech. rep., Duke University.
- Cook, A., Gibson, G., Gilligan, C., 2008. Optimal observation times in experimental epidemic processes. *Biometrics* 64 (3), 860–868.
- D'Argenio, D. Z., 1981. Optimal sampling times for pharmacokinetic experiments. *Journal of Pharmacokinetics and Pharmacodynamics* 9, 739–756.
- Davidian, M., 2009. Non-linear mixed-effects models. In: *Longitudinal Data Analysis*. Chapman & Hall / CRC Press.
- Davidian, M., Gallant, R., 1993. The nonlinear mixed effects model with a smooth random effects density. *Biometrika* 80, 475–488.
- Diggle, P., Liang, K., Zeger, S., 1994. *Analysis of Longitudinal Data*. Clarendon Press, Oxford.
- Drovandi, C. C., Pettitt, A. N., 2013. Bayesian experimental design for models with intractable likelihoods. *Biometrics* 69(4), 937–948.
- Geweke, J., 1989. Bayesian inference in Econometric models using Monte Carlo integration. *Econometrica* 57 (6), 1317–1339.
- Han, C., Chaloner, K., 2004. Bayesian experimental design for nonlinear mixed-effects models with application to HIV dynamics. *Biometrics* 60, 25–33.

- Hashimoto, Y., Sheiner, L., 1991. Designs for population pharmacodynamics: Value of pharmacokinetic data and population analysis. *Pharmacometrics* 19(3), 333–353.
- Jonsson, E., Wade, J., Karlsson, M., 1996. Comparison of some practical sampling strategies for population pharmacokinetic studies. *Journal of Pharmacokinetics and Biopharmaceutics* 24(2), 245–263.
- Lindstrom, M., Bates, D., 1990. Nonlinear mixed effects models for repeated measures data. *Biometrics* 46(3), 673–687.
- McGree, J., Noble, G., Schneiders, F., Dunstan, A., McKinney, A., Boston, R., Sil-  
lence, M., 2012. A Bayesian approach for estimating detection times in horses:  
exploring the pharmacokinetics of a urinary acepromazine metabolite. *Journal of  
Veterinary Pharmacology and Therapeutics*.
- Mentré, F., Mallet, A., Baccar, D., 1997. Optimal design in random-effects regression  
models. *Biometrika* 84 (2), 429–442.
- Müller, P., 1999. Simulation-based optimal design. *Bayesian Statistics* 6, 459–474.
- Palmer, J., Müller, P., 1998. Bayesian optimal design in population models for haema-  
tological data. *Statistics in Medicine* 17, 1613–1622.
- Pronzato, L., Walter, E., 1985. Robust experiment design via stochastic approximation.  
*Mathematical Biosciences* 75 (1), 103–120.
- Pukelsheim, F., 1993. *Optimal Design of Experiments*. Wiley, New York.
- Rabe-Hesketh, S., Skrondal, A., Pickles, A., 2004. Generalised multilevel structural  
equation modelling. *Psychometrika* 69(2), 167–190.
- Racine-Poon, A., 1985. A Bayesian approach to nonlinear random effects models. *Bio-  
metrics* 41(4), 1015–1023.
- Retout, S., Mentré, F., 2003. Further developments of the Fisher Information Matrix  
in nonlinear mixed effects models with evaluation in population pharmacokinetics.  
*Journal of Biopharmaceutical Statistics* 13 (2), 209–227.

- Ryan, E., Drovandi, C. C., Pettitt, A. N., 2015. Fully Bayesian experimental design for pharmacokinetic studies. *Entropy* 17, 1063–1089.
- Ryan, E., Drovandi, C. C., Thompson, M., Pettitt, A. N., 2014. Towards Bayesian experimental design for nonlinear models that require a large number of sampling times. *Computational Statistics and Data Analysis* 70, 45–60.
- Sheiner, L., Beal, S., 1983. Evaluation of methods for estimating population pharmacokinetic parameters. III. Monoexponential model: Routine clinical pharmacokinetic data. *Journal of Pharmacokinetics and Biopharmaceutics* 11(3), 303–319.
- Sorenson, D., Gianola, D., 2002. Likelihood, Bayesian and MCMC methods in quantitative genetics. Springer, New York.
- Spiegelhalter, D., Freedman, L., Parmar, M., 1996. Bayesian approaches to randomize trials. In: Berry, D., Stangl, D. (Eds.), *Bayesian Biostatistics*. Dekker, New York, pp. 67–108.
- Stigler, S., 1971. Optimal experimental designs for polynomial regression. *Journal of American Statistical Association* 66, 311–318.
- Stroud, J., Müller, P., Rosner, G., 2001. Optimal sampling times in population pharmacokinetic studies. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 50 (3), 345–359.
- Tan, F., Berger, M., 1999. Optimal allocation of time points for the random effects models. *Communications in Statistics* 28(2), 517–540.
- Wakefield, J., 1994. An expected loss approach to the design of dosage regimens via sampling-based methods. *Journal of the Royal Statistical Society. Series D (The Statistician)* 43 (1), 13–29.
- Wolfinger, R., 1993. Laplace’s approximation for nonlinear mixed models. *Biometrika* 80, 791–795.